Gestural delay and gestural reduction. Articulatory variation in /l/-vocalisation in Southern British English

Patrycja Strycharczuk and James M. Scobbie

METADATA

In addition to this metadata, what follows is a formatted version of our AAM (2018), minus figures. Tables appear after the main text. The complete Version of Record is:

Strycharczuk, Patrycja and Scobbie, James M. (2020) Gestural delay and gestural reduction. Articulatory variation in /l/-vocalisation in Southern British English. In Przewozny, Anne, Cécile Viollain and Sylvain Navarro (Eds). **The Corpus Phonology of English: Multifocal Analyses of Variation**, pp. 9-29. Edinburgh University Press.

https://edinburghuniversitypress.com/book-the-corpus-phonology-of-english.html

ISBN
Hardback: 9781474466998
eBook (ePub): 9781474467025
eBook (PDF): 9781474467018

Abstract

The vocalisation of /l/, as currently observed in Southern British English (SBE), involves weakening of the consonantal tongue tip (TT) gesture. Such weakening can be conceptualised in terms of spatial reduction, where the magnitude of the TT gesture is decreased, or in terms of temporal delay, where the tongue tip gesture occurs relatively late, sometimes becoming masked. In this paper, we use a corpus of articulatory (ultrasound) data to tease apart the relative contribution of delay and reduction in ongoing /l/-vocalisation in SBE. The most extreme case of vocalisation we observe involves deletion of the TT gesture. More frequently, we find gradient reduction in gestural magnitude, which may be accompanied by gestural delay. For one of our speakers, the TT gesture is delayed to the point of becoming covert. However, the considerable delay observed in this case is proportional to the advanced degree of gestural reduction. We argue for an interpretation where /l/-vocalisation is primarily a spatial phenomenon, and delay is mostly a secondary manifestation of weakening. We consider the significance of our findings to more abstractionist approaches, and their view of /l/-vocalisation as a categorical phenomenon.

Gestural delay and gestural reduction. Articulatory variation in /l/-vocalisation in Southern British English

Patrycja Strycharczuk and James M. Scobbie

1 Introduction

The vocalisation of /l/ in Southern British English (SBE) is an example of a weakening change, in which the consonantal tongue tip (TT) occlusion is lost. Such loss is assumed to be phonetically gradual, and can be straightforwardly modelled in spatial terms as an incremental reduction of the degree of constriction. However, the loss of constriction can also be modelled in terms of articulatory delay, where the tongue tip gesture is present, but it occurs relatively late, such that acoustic consequences of tongue tip contact are not realised. The role of gestural delay in conditioning /l/-vocalisation is acknowledged by previous studies (e.g. Browman & Goldstein 1992; Gick 1999; Tollfree 1999), but it is not fully understood. We do not know whether gestural delay is an independent mechanism in /l/-vocalisation, or whether it is strictly tied to specific degrees of gestural reduction. We address this issue in the present paper, looking at the relative contribution of delay and reduction in advanced and incipient /l/-vocalisation in SBE, using ultrasound data from a previously collected corpus (Strycharczuk & Scobbie 2015; Strycharczuk & Scobbie 2016).

The corpus we use here was not compiled with /l/-vocalisation in mind, but it is a suitable resource for a study of this phenomenon, since it contains articulatory data on /l/ in systematically-controlled segmental and prosodic position, including word-final pre-consonantal/pre-pausal /l/, i.e. the potential context for vocalisation. The motivation for the present study comes from our earlier casual observation that some speakers in the corpus clearly vocalise their word-final /l/, but they achieve vocalisation in different ways. On the one hand, we see varying degrees of gestural reduction in tongue tip raising, from complete alveolar contact to no raising at all. On the other hand, we also find instances where speakers raise the tongue tip when pronouncing a word-final /l/, but the gesture is delayed beyond the offset of voicing, and consequently, the gesture is not fully audible. This type of covert articulation has previously been observed in studies of word-final /r/-weakening in Standard Dutch (Sebregts 2015) and Glasgow English (Lawson, Stuart-Smith & Scobbie 2018), and it is also mentioned as an articulatory strategy involved in the production of /l/-vocalisation in a single speaker of American English (Recasens & Farnetani 1994). In this context, further documentation of such extreme delay in /l/-vocalisation is one of the goals of the present paper. Another, related, goal is to illuminate the relative contribution of gestural reduction and gestural delay in incipient /l/-vocalisation. In addition to categorical vocalisers in our corpus, we also find speakers who clearly produce a TT gesture in word-final /l/, although the magnitude and timing of the gesture may vary. We analyse data from these speakers, in order to verify whether gestural magnitude and timing are gradiently reduced, and if so, how such reduction relates to the more categorical patterns we find in advanced vocalisers.

1.1 /l/-vocalisation in Southern British English

/l/-vocalisation has been noted in Southern British English for at least three decades. The first reports of /l/-vocalisation identify it as a feature of London English (Gimson 1980, 202–203). Wells (1982) also observes that /l/-vocalisation is variable, and thus it should be considered a change in progress. Subsequent reports show that /l/-vocalisation spread to other regions of the UK, including Cambridge (Wright 1987; Wright 1988), Essex and the Fens (Johnson & Britain 2007), and it is also found in Scottish English (Scobbie, Pouplier & Wrench 2007; Scobbie & Pouplier 2010; Stuart-Smith et al. 2013). The studies find a degree of variation, conditioned by phonological environment (Wells 1982; Wright 1987; Hardcastle & Barry 1989), style (Wright 1988), and socio-linguistic factors, including social practices (Stuart-Smith et al. 2013). Notably, numerous authors also comment on the variation concerning the degree of tongue tip weakening in /l/-vocalisation. Two early electropalatography (EPG) studies on /l/-vocalisation, Wright (1987) and Hardcastle & Barry (1989), note instances of partial loss of alveolar contact in coda /l/, where there is some contact between the tongue tip and the alveolar ridge. These cases exist alongside apparently more categorical variants, in which the tongue tip contact is lost altogether. Hardcastle & Barry (1989) acknowledge this as a methodological challenge that concerns classifying individual tokens of /l/ as vocalised or not. They go on to argue that instances of partial alveolar contact support the interpretation of /l/-vocalisation as an inherently gradient phenomenon. The hypothesis here is that suppression of alveolar contact is not a primary factor in vocalisation. Instead, the gesture is weakened, which leads to segmental variation: there may be no alveolar contact whatsoever before dorsals (e.g. *milk*), whereas central alveolar contact is retained before other alveolar targets (e.g. *built*), due to coarticulation. Scobbie & Pouplier (2010) also find both gradient and complete reduction of alveolar contact, which is mainly speaker dependent. Different speakers (5 speakers of SBE and 5 speakers of Scottish English) produce varying degrees of alveolar contact in coda /l/ in the same segmental context (immediately before word-initial/b/ or /h/ in the following word). Some speakers have no alveolar contact in this context. However, even within speakers who do make the contact, the degree of contact varies. This behaviour appears to be the continuation of the change described by Hardcastle & Barry (1989): gradient reduction in the degree of alveolar contact is conditioned by gradient reduction in the magnitude of the apical gesture, and the gesture continues to reduce even after the contact is lost.

The speakers who have no alveolar contact appear to be, in some sense, categorically distinct from speakers who make alveolar contact, even if it is weakened. However, the absence of alveolar contact does not entail the absence of a tongue tip *gesture* in /l/-vocalisers. Indeed, we know of cases where there is gradient variation affecting the tongue tip position in /l/ in speakers who are vocalisers (i.e. they do not make alveolar contact in coda /l/). Some such cases are described by Wrench & Scobbie (2003), who analyse /l/-vocalisation using electromagnetic articulography (EMA) data and ultrasound. One of the measures they use is the distance of the tongue tip sensor from the palate. This distance tends to vary gradiently: it is greater for pre-pausal /l/ compared to pre-labial. For pre-vocalic /l/, the distance is smaller yet, although it is still different from non-vocalised word-initial /l/s, where the sensor makes contact with the palate, so that average distance is zero. A similar gradient pattern of tongue tip reduction can be seen in ultrasound data from an Essex speaker presented in Turton (2014). Turton compares averaged tongue tracings in /l/ in a variety of contexts stratified according to prosodic and morphological factors. She finds patterns of /l/-vocalisation in one speaker from Essex and two speakers from London. Data from these three speakers confirm that the tongue tip is consistently down for all word-final contexts. However, for the Essex speaker, the degree of tongue tip raising appears to form a continuum from word-final *heal* (where the tongue tip is relatively lowest) to word-medial *healing* (where the tongue tip is raised, and there is no audible vocalisation). In contrast, there are more robust differences between *heal*, *heal it* and *healing* in the tongue root and tongue dorsum. The two London speakers in Turton's study show more of a categorical break in the degree of tongue tip raising between vocalised word-final /l/s and non-vocalised word-medial and word-initial ones.

Turton (2014) discusses /l/-vocalisation in the context of another change affecting /l/, namely /l/-darkening. In many English dialects coda /l/s are dark, which can be manifested by dorsal retraction or velarisation. Thus, the vocalic gesture that forms a part of /l/ becomes augmented. The tongue tip raising is sometimes talked about as the consonantal /l/-gesture (e.g. Gick (1999), see also Section 1.2 below). Weakening of this consonantal gesture could be viewed as an extension to increased dorsalisation, though the articulatory mechanisms for this are poorly understood. The enhancement of vocalic features and suppression of consonantal ones lead to /l/ being more vowel-like (see Section 4 for more discussion on the relationship between /l/-darkening and /l/-vocalisation).

1.2 Temporal factors in /l/-vocalisation

The individual studies surveyed in Section 1.1 are relatively small in scope and they use a variety of test items and methods, but nevertheless, a coherent picture of gradient spatial reduction emerges when the various pieces of evidence are put together. We find instances of gradient reduction of the degree of apical contact in some speakers with incipient /l/-vocalisation, whereas in speakers with more advanced vocalisation, we may find gradient reduction pattern of the magnitude of the tongue tip gesture. These two can be seen as subsequent steps of a single reduction pathway: initial less advanced reduction of the TT gesture may, under certain circumstances, result in partial alveolar contact. Once the reduction progresses, alveolar contact is no longer found, but we still see ongoing reduction in TT magnitude, depending on the phonological environment. However, such a picture is not complete, because it does not take into account temporal aspects of /l/-articulation and /l/-reduction.

To an extent, the opposition between dark and vocalised /l/ can be modelled just in terms of gestural magnitude: /l/-darkening involves increased dorsal retraction. However, going back to the seminal paper by Sproat & Fujimura (1993), timing has also been considered a key component in modelling the articulation of /l/. Sproat & Fujimura (1993) identify two distinct gestures that make an /l/: a dorsal gesture (retraction and/or raising of the back part of the tongue), and an apical gesture (raising of the tongue tip).[1] In an X-ray microbeam study of American English, Sproat & Fujimura (1993) find that both dark and light /l/ contain a dorsal and an apical gesture. However, the detail of the sequence of these two gestures can vary. In dark /l/, the apical gesture is delayed relative to the dorsal one. Furthermore, the tip delay increases gradiently depending on the strength of the levels of morphosyntactic boundary following /l/. There is increasing tip delay, comparing *Mr Beelik*, *beal-ing*, *Beel equates* and *Beel*. This observation by Sproat and Fujimura represents a major shift from an earlier, static, conceptualisation of /l/-positional variation (e.g. Giles & Moll (1975), and it shows that it is crucial to consider dynamic aspects of /l/-articulation in a description of darkening. In fact, Sproat and Fujimura argue that gestural magnitude is a consequence of position-specific gestural timing. Specifically, the dorsal gesture's magnitude can only be realised fully in word-final rimes, which are typically longer. In the relatively shorter word-initial and word-medial positions, on the other hand, the dorsal gesture is subject to articulatory undershoot.

This dynamic model also opens up the possibility that tip delay is one of the mechanisms involved in the actuation of /l/-vocalisation. If /l/-darkening involves a relatively late tongue tip gesture, we can also imagine a situation where the TT gesture overlaps with the subsequent gesture (e.g. with the labial closure in *feel me* across words or more commonly within such monomorphemic words as *album*, *elbow* or *alpha*, or complex ones like *almost* and *always*). In such case, the TT gesture may be present, but is articulatorily masked and so may be inaudible, giving the percept of /l/-vocalisation. This hypothesis is formulated by Tollfree (1999), based on the models by Sproat & Fujimura (1993) and by Browman & Goldstein (1992), though only for pre-consonantal /l/. According, to Tollfree, in word-final pre-pausal /l/, the tip gesture should always be audible. However, we can imagine a different scenario, in which gestural delay becomes phonologised, and increasing levels of such delay cause vocalisation even pre-pausally. A potentially attractive aspect of the delay-driven view of /l/-vocalisation is that /l/-vocalisation is a straightforward extension of tip delay, a phenomenon independently observed in /l/-darkening.

[1]Recasens (2016) argues that the two gestures are not identifiable in systems where dark /l/ is phonemic. In Southern British English, both gestures are measurable, at least in some vocalic contexts (Strycharczuk & Scobbie 2015).

Another possible way to think about gestural delay is that delay is one of the components in a more holistic process of weakening. This account seems consistent with a model put forward by Gick (1999) for final vocalisation of liquids. Gick proposes that vocalisation of final /r/ and /l/ inherently consist in the weakening of the consonantal gesture in coda, which is "both reduced in magnitude (final reduction) and temporally delayed". In this account, gestural delay is not so much a driving force for vocalisation, as one of its two closely correlated manifestations, along with spatial reduction.

1.3 This study

In some of our own recent work, we analysed the timing of the apical gesture relative to the dorsal one for /l/ in SBE (Strycharczuk & Scobbie 2015). For word-final /l/, as produced by 6 native speakers of SBE, we analysed the temporal relationship between the dorsal and the apical gesture, and concluded that the apical gesture was indeed delayed relative to the dorsal one. According to our measure, the delay was ca. 120ms. Furthermore, there was an apparent-time effect, whereby younger speakers had greater tip delay than older speakers. The apparent-time effect was as much as 75ms in its most prominent context, namely when word-final /l/ was preceded by /uː/. Nevertheless, the TT maximum still occurred before the offset of voicing for all the 6 speakers included in our 2015 study.

Here we extend our previous work to also investigate how the observed delay of the TT gesture relates to gestural magnitude. We provide such an analysis for 11 speakers, using a modified measure of magnitude as developed by Lin, Beddor & Coetzee (2014). We are also concerned with the acoustic consequences of the previously observed tip delay. We assess this in two ways. Firstly, we measure the time lag between the apical gesture and the offset of voicing in final /l/ as a measure of the extent to which TT gesture is audible. We then investigate the acoustic effect of TT gestures depending on how late they are, relative to the voicing offset. The spectral measure we focus on is F2−F1 distance, an important correlate of /l/-darkening/vocalisation (Sproat & Fujimura 1993; Lin, Beddor & Coetzee 2014; Turton 2014).

2 Materials and methods

2.1 Materials

The speech materials we analyse come from the corpus on segmental and morphological effects on SSBE /uː/ and /ʊ/-fronting (Strycharczuk & Scobbie 2016; Strycharczuk & Scobbie 2017). For the present study, we selected 864 tokens of /l/ in three contexts, where we expect varying degrees of /l/-darkening or /l/-vocalisation: i) word-final pre-consonantal (*fool#five*); ii) word-final pre-vocalic (*fool#it*); and iii) word-medial (*fooling*). The preceding vowels were always /uː/ or

/ʊ/. The word-final pre-consonantal context (*fool#five*) is the potential environment for vocalisation, whereas the word-medial context (*fooling*) is a baseline in which no vocalisation is expected. In the word-final pre-vocalic context, we expect some gradient weakening, but less so than in the word-final pre-consonantal context.

2.2 Speakers

For the purpose of the present study we analysed data from 11 speakers, 6 younger (under 25) and 5 older (45+). All the speakers came from and grew up in the South of England, although they all spent some time in Scotland prior to the experiment. All speakers were female, except one older male. Six of the speakers had been included in our previous study on tip delay (Strycharczuk & Scobbie 2015). The additional five speakers were selected from the corpus based on the quality of their ultrasound image. We chose all speakers for whom we could consistently track the tongue contour in the anterior part of the tongue.

2.3 Acoustic analysis

For the acoustic analysis, we manually annotated the duration of the vowel and the following /l/, and we measured the first three formants at 10% intervals throughout the vowel+/l/ sequence, using a Praat script modified from Remijsen (2004). In this study, we use the formant measurement towards the end of the /Vl/ sequence (90% into its duration), since we are interested in the acoustic consequences of articulatory events in the later part of the sequence where the phonetic reflexes of /l/ are most evident.

2.4 Articulatory analysis

The crucial aspect of our articulatory analysis is identifying the tongue tip gesture, and quantifying it in terms of magnitude and timing. The first step in this process was tracking tongue contour in the ultrasound data for the entire /Vl/ sequence, as well as throughout the preceding and following segment (defined acoustically). The contour tracking was performed semi-automatically by the Articulate Assistant Advanced software (AAA2.16 Articulate Instruments Ltd., 2014), supervised by the first author. Manual corrections were performed where necessary.

The tongue contour tracks were then used to identify the tongue tip gesture. We did this based on maximum displacement along a specific fan line chosen to best represent the approximation of the tongue to the alveolar ridge as a linear variable, analysed as it changed in normalized magnitude (see below). The fan lines are 42 equidistant radials, with a maximal angle of ±67° relative to the centre of the ultrasound probe. Together they make up a reference frame (a fan) in which tongue shape and location can be plotted, and changes in these parameters extracted for regions of interest in the vocal tract.

These speaker-dependent regions of interest were determined based on the pattern and extent of tongue surface movement through the region, with an eye to the specific radial vector on which the changing constriction would be measured. There is no a-priori linear dimension (such as 'vertical') which is a preferred one for quantifying the approximation of one three-dimensional tongue body to another. This issue persists for the 2D surface tracings we use. Therefore, for each speaker, we inspected 10 example tokens by exporting all the tongue tracks from each token into a single coordinate frame. We then identified an anterior fan line along which the relatively greatest tongue displacement occurred, as illustrated in Figure 1. This was done by visual inspection, combined with examination of means and standard deviations of displacement values throughout a token for each fan line (Strycharczuk & Scobbie 2015; Heyde et al. 2016). We picked fan lines for which the standard deviation was relatively highest (indicating the most movement), but also paying attention to confidence values, since tongue contour tracking in the tongue tip area is not always reliable.

INSERT FIGURE 1 APROXIMATELY HERE

Figure 1 Example measurement of aperture for a selected token. The individual splines represent tongue contour tracings throughout the latter portion of the vowel and throughout the /l/. The hard palate contour is traced in green. Tongue tip is on the right. The aperture measurement (in blue) is the distance along the fan line with relative greatest variation in tongue displacement in the alveolar region. Aperture is measured between the spline representing maximum displacement and the hard palate.

We then analysed the displacement of the tongue contour tracing along the selected fan line and picked the contour with maximum displacement as representative of the TT maximum. For all TT maxima, we measured the distance between tongue contour and the hard palate along the selected fan line (see Figure 1). This approach is inspired by Lin, Beddor & Coetzee (2014), who also measured the distance between tongue tip and the hard palate tracing, although the distance was defined with respect to the image's vertical orientation. We relate to the distance measure as 'aperture'. We z-score normalised the aperture measure within speaker, since for some speakers the tongue tip and the hard palate are imaged better than for others. Normalisation addresses this issue to a certain extent, since we included a within-speaker baseline, i.e. the morpheme-final context (*fool-ing*), where we expect full tongue tip contact.

We also recorded the time point at which the maximum occurred, and measured the time lag between this point and the acoustic offset of voicing in /l/ (for the *fool#five* context only). The greater the time lag from TT maximum to the offset of voicing, the bigger the relative delay, and the greater likelihood of a percept of vocalisation, all things being equal. A relatively earlier TT maximum (less delay) will appear less vocalised. We also wanted to know whether there is a continuum of relative timing between tongue tip raising and the offset of voicing.

For one of the speakers who is an extreme vocaliser, YF6, there was hardly any tongue tip displacement in the word-final pre-consonantal context (*fool#five*), and in consequence, the TT gesture was not identifiable (note that Scobbie & Pouplier 2010 report a similar issue). For this speaker, we took the final ultrasound frame in the region corresponding to the acoustic duration of /Vl/ as representative of the maximum. In Section 3.2, we show data from this speakers as an illustration of extreme reduction. Time lag was not measurable for this speaker.

We complement the analysis of aperture and lag with analysis of averaged tongue contours (per speaker and per vowel) at the tongue tip maximum. We used Smoothing Spline Analysis of Variance (SS-ANOVA) to obtain the averaged smooths (Davidson 2006; Gu 2013; Gu 2014). We ran the SS-ANOVA in Cartesian coordinates, as opposed to polar, since we find that the analysis using polar transformation, as it was operationalised, sometimes distorted the tongue tip.

3 Statistical analysis and results

3.1 TT aperture as a function of context

As the first step in our analysis, we investigated how aperture depends on the context, and how this changes depending on speaker age. We fitted a linear mixed-effects regression model with aperture as the dependent variable, and with context and age as main effects. We then added a random slope for context within speaker, and found that the fit of the model improved significantly ($\chi^2 = 57.00$, $\mathrm{df} = 5$, $p < 0.001$). This suggests considerable individual variation with respect to aperture. In contrast, there was no significant main effect of age, as removing this predictor did not significantly affect the model fit ($\chi^2 = 0.03$, $\mathrm{df} = 1$, $p = 0.86$). There was no significant interaction between age and context, as established by adding the interaction and comparing the model to the one with main effects only ($\chi^2 = 1.27$, $\mathrm{df} = 2$, $p = 0.53$). The final model we present in Table 1 has age and context as main predictors, random intercepts for speaker and item, and a random slope for context within speaker. The p-values in this and subsequent model summaries were obtained using the stargazer package (Hlavac 2018), which also served as the model for formatting the model summaries in this paper.

INSERT TABLE 1 APROXIMATELY HERE

As shown in Table 1, the average aperture is significantly greater in the word-final pre-consonantal condition (*fool#five*), but there is no significant difference in average aperture between the intercept (word-medial pre-vocalic) and the word-final pre-vocalic condition. The effect of age is not significant.

The results of the modelling confirm a degree of /l/-vocalisation in the word-final pre-consonantal position, but not in the word-final pre-vocalic context. Impressionistically, there were instances of word-final pre-vocalic vocalisation, but they were rare, and there was no significant difference in the gradient reduction in aperture in this context. We do not observe any significant apparent time effects and no interactions, which suggests that younger speakers do not have more vocalisation in the context where it is most expected, i.e. word-finally before a consonant. However, the absence of a significant main effect seems to be due to inter-speaker variability, as would be expected in a situation of variation and change. This is confirmed by the analysis of by-speaker coefficients showing the effect of context within participant illustrated in Figure 2.

As shown in Figure 2, there seems to be an overall increase in aperture in the word-final pre-V context compared to word-medial, but speakers vary in how much increase in aperture they show. The word-final pre-C context involves the most inter-speaker variation. The most extreme vocalisers, YF6, YF4 and YF5, are in the younger speaker group. At the same time, other younger speakers, speakers YF1 and YF7 show no relative increase in aperture (no /l/-vocalisation). Older speakers seem to be relatively more consistent. Another noteworthy observation is that average aperture values in the word-final pre-C condition are fairly continuous: from speaker YF7 (who has no /l/-vocalisation) to speaker YF6 (who is an extreme vocaliser with no residue of the TT gesture): all intermediate degrees of aperture are attested. In Section 3.2 below, we analyse average tongue contours from selected speakers, the ones labelled in Figure 2, to provide a more holistic illustration of entire tongue contours in vocalising and non-vocalising contexts.

INSERT FIGURE 2 APPROXIMATELY HERE

Figure 2 Average aperture depending on context and speaker. Younger speakers are in blue, and older speakers are in red. Data from the labelled speakers are analysed in more detail in Section 3.2

3.2 Degrees of /l/-vocalisation

Figure 3 shows averaged tongue contour tracings for selected speakers, arranged from most extreme vocaliser, YF6, to YF7. YF6 had no TT gesture in the word-final pre-consonantal position, such that we could not measure it (for this speaker, tongue contours were extracted at the acoustic offset of /l/). The averaged smooths clearly show that the tongue tip is down for this speaker in the word-final pre-consonantal position, whereas there is tongue tip raising in the word-medial context, as well as in the word-final pre-vocalic context. Also note that this speaker seems to show some gradient reduction in the TT contact for the word-final pre-vocalic context when the preceding vowel is /ʊ/.

INSERT FIGURE 3 APPROXIMATELY HERE

Figure 3 Comparison of averaged tongue contours for selected speakers, depending on context and the vowel preceding /l/

Another extreme vocaliser is speaker YF4. Unlike YF6, YF4 did produce some TT raising in the word-final pre-consonantal condition, but we judge that there was never any TT contact in this context. What is particularly striking about this speaker is that word-finally before a consonant, the TT maximum occurred very late indeed. On average, the TT maximum was reached 50ms after the voicing offset. We show an example token where such delay occurred in Figure 4. The tongue dorsum was not very well imaged for the speaker, but we can see the anterior part of the tongue relatively clearly. Comparing the two frames, the tongue blade continues moving upwards following the offset of voicing, and reaches its maximum after the voicing had ceased. Thus, although the TT gesture is present for this speaker, its severe reduction in magnitude and its considerable delay both contribute to an extreme acoustic reduction, typical of the type that renders any alveolar gesture of the /l/ acoustically untraceable, or covert.

INSERT FIGURE 4 APPROXIMATELY HERE

Figure 4 Example of gestural delay by speaker YF4. The left panel shows the image the tongue at the offset of acoustic voicing. The right panel shows the image of the tongue when it reaches maximum TT displacement, ca. 50 ms after the offset of voicing.

Speaker YF5 is also a vocaliser who appears to have no tongue tip contact in the word-final pre-C context when the preceding vowel is /uː/, although there seems to be contact word-finally before a consonant when the preceding vowel is /ʊ/. Unlike YF6, however, YF5 shows some tongue tip raising, and the TT gesture is not delayed beyond acoustic offset (as it happens for YF4). Thus, YF5's vocalisation seems more gradient, compared to either YF6 or YF4.

Another gradient vocaliser is speaker YF8, and once again, the gradience is more apparent in the /uː/ context. Here we see that the tongue tip is raised in the word-final pre-consonantal context (as it is in the baseline context), but there seems to be a trend towards gradient reduction in the degree of tongue tip contact.

At the other end of the spectrum, we find speakers like YF7, who has the same degree of TT contact across all three contexts.

In summary, comparison of averaged tongue contours (for selected speakers) on an individual basis is compatible with the results of mixed-effects modelling of aperture reported in Section 3.1 above. For all speakers we find TT raising in the word-medial context, and typically also in the word-final pre-V context. It is in the word-final pre-C context, on the other hand, where we see the biggest potential for structured variation and change, given the wide range of productions observed. Our results quantify the wide range observed, covering all degrees of intermediate TT raising from no raising at all in speaker YF6 to full contact in speaker YF7. In addition, we have observed that speaker YF4 (but no other speaker) delays the maximum TT raising beyond the offset of voicing, such that the TT gesture is potentially inaudible (or its acoustic consequences are reduced). In Section 3.3 below, we turn to analysing the timing of the TT gesture for all speakers, to verify whether intermediate degrees of gradient TT reduction are correlated with degrees of gestural delay.

3.3 Relationship between magnitude and delay in the vocalisation context

In order to analyse the relationship between aperture, timing and acoustics, we fitted a series of linear mixed models. These models were fitted to the data from word-final pre-consonantal items, since it is the environment in which between and within-speaker variation in the degree of vocalisation is potentially greatest. We excluded speaker YF6 from this part of the analysis, because she had complete deletion of the TT gesture, which rendered acoustic delay not measurable.

We fitted a model predicting aperture based on gestural lag, vowel, speaker age and the F2—F1 difference. The F2—F1 difference was centred to improve model convergence. We then tested for all possible interactions and for random slopes, but found no significant improvements in model fit. The best-fitting model we present in Table 2 had only main effects, as well as random intercepts for speaker and item.

INSERT TABLE 2 APPROXIMATELY HERE

Aperture increases significantly with gestural delay: delayed gestures are more likely to also be reduced. We also find, in line with previous reports, that decreased gestural magnitude is correlated with F1 and F2 coming together. There is no significant effect of preceding vowel, or speaker age. The effect of gestural delay on aperture is illustrated in Figure 5. The figure shows the raw data, and the line of best fit, according to the model. We can see that early TT gestures generally have small aperture values, i.e. they are fully realised, whereas considerably delayed gestures are invariably reduced. There is, however, also a number of intermediate data points that correspond to gestures that are close to the acoustic offset of voicing, but where the maximum is reached before the voicing ceases. For this group, there is a general correlation between delay and increase in aperture, but we also find variation in this band: some gestures achieved ca. 50ms before voicing offset are fully realised, whereas others may be reduced.

INSERT FIGURE 5 APPROXIMATELY HERE

Figure 5 The effect of gestural delay on aperture

The next model we fitted predicted gestural delay based on vowel preceding /l/, speaker age, aperture and F2−F1 distance. We then tested for interactions and random slopes, and found a significant model improvement after including an interaction between vowel and speaker age (χ^2 =16.25, df =1, $p < 0.01$). Furthermore, the model improved significantly when we also added an additional effect, that of combined vowel + /l/ duration (χ^2 =73.04, df =1, $p < 0.01$). The best-fitting model, whose fixed part is summarised in Table 3, had main effects of F2−F1 distance aperture, /Vl/ duration and an interaction between vowel and speaker age.

The only significant main effect is /Vl/ duration. There is a very strong negative correlation between duration and delay: in shorter segments, the TT gesture tends to occur relatively closer to the offset of voicing. In addition, we find that gestural delay is relatively greater for /uː/ than for /ʊ/ in older speakers, whereas in younger speakers, delay is comparable for both vowels. Neither aperture nor F2−F1 distance are significant predictors in the delay model. The absence of a significant effect of F2−F1 is perhaps not surprising, considering that we expect the greatest effect of delay on formants when the TT gesture is delayed beyond the offset of the sound source, and that only happens for one speaker in our data (YF4). The absence of an effect of aperture is more surprising, given that delay and aperture are strongly correlated in our previous model (Table 2 and Figure 5).

INSERT TABLE 3 APPROXIMATELY HERE

3.4 Results summary

Putting together different types of analysis presented in this section, the following findings emerge. There are varying degrees of /l/-vocalisation (operationalised as the degree of aperture, a speaker-specific linear, radial distance between some part of the tongue surface and part of the hard palate in the alveolar region of interest). Along this dimension, a whole spectrum of degrees of vocalisation is attested: from a full gesture resulting in contact between the tongue tip and the palate to complete deletion of the tongue tip gesture. Individual speakers may occupy any space along this continuum, i.e. varying gradient degrees of reduction in TT magnitude are found. We do not find significant main speaker age effects, but analysis of individual variation shows that this is because younger speakers are extremely varied (from extreme vocalisation in some to none at all in others), whereas older speakers generally have TT contact (either full or partially reduced). Vocalisation is generally confined to word-final pre-consonantal context (out of the contexts we tested). We find very little of it in the word-final pre-vocalic context.

Regarding the relationship between increase in aperture and gestural delay, we find some correlation, but there are also caveats. Early TT gestures are always fully realised, while the extremely delayed ones (past the offset of voicing) are invariably reduced. However, somewhat late but still fully audible TT gestures may or may not be reduced: there is some correlation between aperture and delay in this group, but there is also big variance. We also find that gradient gestural delay does not result in F2 and F1 coming together in the way gradient increase in aperture does. Finally, there is a very strong correlation between timing of the TT gesture and the duration of the /Vl/ sequence: the longer the sequence, the greater the delay of the TT maximum relative to the acoustic offset.

4 Discussion

In Section 1, we introduced two conceptualisations of /l/-vocalisation. According to one, /l/-vocalisation is primarily a gradual reduction in the *magnitude* of the TT gesture. The alternative proposal is that /l/-vocalisation is mainly driven by gestural *timing* (specifically, a delay). If the TT gesture is frequently delayed, and consequently masked by the following gesture (or by the absence of sound source), this may lead to a perception-driven change, where listeners start weakening the consonantal gesture in their own speech. In principle, both scenarios of change may lead to the same result: considerable reduction accompanied by considerable delay (as seen in Speaker YF4 in our data), and eventually, deletion of the TT gestures (as seen in speaker YF6). However, even if the end result is the same, discontinuities in the data and inter-speaker differences can give us clues about how transmission of vocalisation is achieved. If vocalisation is understood as gradual reduction of the magnitude of the TT gestures, this should be reflected in gradient variation of the TT aperture within speaker and across speakers. If, on the other hand, gestural reduction is a secondary diachronic consequence of delay, we may expect to see discontinuities in the degree of reduction and a more gradient distribution of the degree of delay.

Overall, our data appear more consistent with the idea that /l/-vocalisation is primarily about reduction in magnitude. The main argument for this comes from the observation that virtually all degrees of aperture are attested in our sample: from full contact, via intermediate degrees of increase in aperture, to full gestural reduction. While we find all possible degrees of reduction in TT magnitude, delay appears to be a secondary factor. All of our speakers who have any degree of vocalisation use TT reduction, but only one of them consistently uses gestural delay, and for this speaker, the degree of delay is proportional to the reduction in TT magnitude.

In contrast, we do not find instances where gestural delay occurs without gestural reduction. Such a possibility is predicted under the ‘delay-first’ scenario, where gestural reduction first appears due to perceptual reinterpretation of reduction as delay. Of course, the absence of such a pattern in our sample does not preclude its presence in a larger speaker population or different dialect. An additional consideration is that this type of delay is likely to be a property of casual speech that might be difficult to elicit under the conditions of instrumental research. Thus, while our current results seem more in line with the 'reduction-first' proposal, this interpretation would be considerably strengthened if we had evidence from other speakers who also make covert TT gestures in their final /l/. Data from such speakers could give us a better understanding of whether delay is always a factor involved in more advanced instances of vocalisation before the tongue tip gesture undergoes deletion.

In general, our findings invite more research into all speakers who appear to have /l/-vocalisation in English. Speakers who have lost the TT contact are by no means a monolithic group. On the contrary, there is variation among them which is strongly indicative of continuing sound change. We can generalise that gradient reduction of tongue tip contact (as shown in earlier EPG studies, e.g. Hardcastle & Barry 1989) continues as a gradient reduction of the tongue tip gesture. What has yet to be investigated is how reduction of the TT gesture propagates through speech communities once the TT contact has been lost.

Our results also challenge the view that /l/-vocalisation is a form of 'extreme darkening'. This kind of interpretation could be pursued if we take a strictly temporal interpretation of /l/-vocalisation, where /l/-darkening involves relative delay of the apical gesture beyond the dorsal one, and vocalisation is a step further, involving gestural delay that makes the apical gesture inaudible. Such a view is challenged by speakers who make a very strong use of spatial reduction in producing vocalised /l/, most notably Speaker YF6. We observed in Section 2.4, that we could not identify a distinct TT gesture for this speaker in word-final pre-C /l/s, and so we were unable to measure gestural delay for this speaker. It is problematic to define vocalisation in terms of delay when such delay is not measurable, and so the vocalised variants appear to be, in some sense, categorically different from the tokens of dark /l/. By the same token, gestural deletion complicates measurements of aperture. Although we can still measure the distance between the tongue surface and the hard palate, we may not want to refer to it as a measure of gestural reduction, *if there is no gesture*. The spatio-temporal consequences of gestural deletion preclude talking about it in terms defined for /l/-darkening.

While our data support the idea that there are prototypical and categorically distinct variants of light, dark and vocalised /l/, category boundaries are difficult to establish, which challenges attempts at categorising different variants of /l/ in terms of broadly-defined patterns. Such a categorisation is inherent to a phonemic or segmental analysis, which relies on categorically different allophonic variants of /l/ in formulating phonological generalisations. For instance, the clear /l/ vs. dark /l/ opposition is often analysed in terms of two segmental units whose distribution isconditioned by syllable structure: clear /l/ occurs in onsets, and dark /l/ occurs in codas. /l/-vocalisation seems to be sensitive to the interaction between segmental and prosodic factors: while vocalisation occurs word-finally, it is usually limited when a vowel follows in the next word. However, although such generalisations approximate observed patterns to some extent, they are difficult to reconcile with the demonstrably gradient and continuous nature of articulatory adjustments affecting /l/ in different contexts. The change from a clear /l/ to a dark one involves incremental increase to the dorsal gesture, coupled with relative delay of TT raising. /l/-vocalisation can be characterised as gradual loss of the TT contact, but we also know that the TT gesture may continue to undergo reduction in ostensibly vocalised cases. For those reasons, an allophonic split into three variants, clear, dark and vocalised /l/ necessarily involves arbitrary cuts to an inherently gradient phenomenon, merely replicating categorical structural differences in a segmental domain.

The challenge to forcing different types of /l/ into broad categories even within the same structural position has been acknowledged from the methodological point of view by sociolinguists interested in quantifying the frequency of different /l/-variants in a speech community. The feasibility of identifying and labelling vocalised /l/ is discussed by Hall-Lew & Fix (2012), who note that auditory rating yields relatively consistent results, but confidence drops for more intermediate cases, which are often also the ones that are sociolinguistically interesting. Accumulating results of instrumental articulatory research into /l/-vocalisation suggest that labelling difficulties are not strictly due to the somewhat crude nature of auditory coding. Rather, these issues follow both from the continuous nature of variation between articulatory variants within one context and the difficulties of comparing similar articulatory variants across structurally distinct contexts. Even with much more detailed articulatory evidence at hand it is difficult to classify some intermediate cases.

Furthermore, different cases may be considered intermediate, depending on which phonetic dimension is being measured. If we focus on the degree of alveolar contact as the primary dimension, as is done in EPG studies, we might view partial contact cases as intermediate, and no contact cases as categorical /l/-vocalisation (e.g. Scobbie & Pouplier 2010). However, we know that the weakening of the tongue tip gesture goes further than that, so we may also want to consider the cases where TT gesture is reduced as intermediate (or gradient), and cases where TT gesture is deleted as categorical. Therefore, some apparent categorical breaks may tell us more about the method we used to observe them than about the abstract organisation of the underlying system.

These considerations prompt us to recommend a mixture of different methodologies for further studies of /l/-vocalisation. The method we used, ultrasound tongue imaging, has the advantage of giving us a holistic view of the tongue posture and movement, unlike EPG, which cannot tell us about the tongue position where there is no contact with the palate. However, since the very tip of the tongue is typically not visible in the ultrasound image, this method necessarily involves inferring the tongue tip position based on the data from the tongue blade. EMA can give us more precise measurements of tongue tip displacement and velocity, but it is less well suited to studying the relationship between articulation and acoustics, since articulation is partially impeded in EMA studies by the presence of wires in the speaker's mouth, and coils need to be placed away from the precise location of contact. Furthermore, for practical reasons, like cost and portability, ultrasound is more appropriate for sociophonetic studies. A larger-scale carefully-sampled study of a single vocalising community would be the natural next step, and the best-case scenario would involve cross-comparisons of results from two methods (e.g. ultrasound and EMA) from a subset of the speakers, preferably incorporating within-speaker variation in style or register. Furthermore, for a more complete picture of /l/-vocalisation, labialisation also ought to be considered. For this reason, a fixed-ratio head-mounted video camera could be a useful additional tool in future instrumental studies of this phenomenon beyond the laboratory.

Acknowledgments

We wish to thank the speakers for participating in our study, Steve Cowen for assistance with the recordings, and Alan Wrench for help with the ultrasound system. We thank an anonymous reviewer for their comments. The research reported in this paper was supported by a British Academy Postdoctoral Fellowship PDF/pf130029 to the first author.

Articulate Instruments Ltd. 2014. *Articulate Assistant Advanced ultrasound module user manual, revision 2.16*. Edinburgh, UK: Articulate Instruments Ltd.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49. 155–180.

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America* 120. 407–415.

Gick, Bryan. 1999. A gesture-based account of intrusive consonants in English. *Phonology* 16. 29–54.

Giles, Stephen B. & Kenneth L. Moll. 1975. Cinefluorographic study of selected allophones of English /l/. *Phonetica* 31. 206–227.

Gimson, A.C. 1980. *An introduction to the pronunciation of English by AC Gimson*. 3rd ed. London: Arnold.

Gu, Chong. 2013. *Smoothing spline ANOVA models*. Vol. 297. Springer Science & Business Media.

Gu, Chong. 2014. Smoothing Spline ANOVA Models: R package gss. *Journal of Statistical Software* 58(5). 1–25. http://www.jstatsoft.org/v58/i05/.

Hall-Lew, Lauren & Sonya Fix. 2012. Perceptual coding reliability of (l)-vocalization in casual speech data. *Lingua* 122. 794–809.

Hardcastle, William & William Barry. 1989. Articulatory and perceptual factors in /l/ vocalisations in English. *Journal of the International Phonetic Association* 15. 3–17.

Heyde, Cornelia, Scobbie, James M., Lickley, Robin and Drake, Eleanor K. 2016. How fluent is the fluent speech of people who stutter? A new approach to measuring kinematics with ultrasound. *Clinical Linguistics and Phonetics*. 30 (3-5): 292-312.

Hlavac, Marek. 2018. *stargazer: Well-formatted regression and summary statistics tables*. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI). https://CRAN.R-project.org/package=stargazer.

Johnson, Wyn & David Britain. 2007. L-vocalisation as a natural phenomenon: explorations in sociophonology. *Language Sciences* 29. 294–315.

Lawson, Eleanor, Jane Stuart-Smith & James M. Scobbie. 2018. The role of gesture delay to coda /r/ weakening: an articulatory, auditory and acoustic study. *Journal of the Acoustical Society of America* 143(3). 1646-1657.

Lin, Susan, Patrice Speeter Beddor & Andries W. Coetzee. 2014. Gestural reduction, lexical frequency, and sound change: A study of post-vocalic /l/. *Laboratory Phonology* 5. 9–36.

Recasens, Daniel. 2016. What is and what is not an articulatory gesture in speech production. *Gradus - Revista Brasileira de Fonologia de Laboratório* 1. 23–42. https://gradusjournal.com/index.php/gradus/article/view/2.

Recasens, Daniel & Edda Farnetani. 1994. Spatiotemporal properties of different allophones of /l/: phonological implications. In (eds.) W.U. Dressler, M. Prinzhorn & J.R. Rennison (Eds) *Phonologica 1992*. *Proceedings of the 7th International Phonology Meeting*. Rosenberg & Sellier, Turin, 195-–204

Remijsen, Bert. 2004. Script to measure & check formants. http://www.lel.ed.ac.uk/~bert/msr\&check$_$f1f2$_$indiv$_$interv.psc

Scobbie, James M. & Marianne Pouplier. 2010. The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English /l/. *Journal of Phonetics* 38. 240–259.

Scobbie, James M., Marianne Pouplier & Alan A Wrench. 2007. Conditioning factors in external sandhi: an EPG study of English /l/ vocalisation. In *Proceedings of the 16th International Congress of Phonetic Sciences*. 441–444.

Sebregts, Koen. 2015. *The sociophonetics and phonology of Dutch r*. Vol. 379. Utrecht: LOT.

Sproat, Richard & Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21. 291–311.

Strycharczuk, Patrycja & James M. Scobbie. 2015. Velocity measures in ultrasound data. Gestural timing of post-vocalic /l/ in English. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Paper 0309. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/proceedings.html.

Strycharczuk, Patrycja & James M. Scobbie. 2016. Gradual or abrupt? The phonetic path to morphologisation. *Journal of Phonetics* 59. 76–91.

Strycharczuk, Patrycja & James M. Scobbie. 2017. Fronting of Southern British English high-back vowels in articulation and acoustics. *The Journal of the Acoustical Society of America* 142. 322–331.

Stuart-Smith, Jane, Gwilym Pryce, Claire Timmins & Barrie Gunter. 2013. Television can also be a factor in language change: Evidence from an urban dialect. *Language* 89. 501–536.

Tollfree, Laura. 1999. South East London English: discrete versus continuous modelling of consonantal reduction. In P. Foulkes and G. Docherty (Eds.) *Urban voices: Accent studies in the British Isles*, 163–184. London: Arnold.

Turton, Danielle. 2014. *Variation in English /l/: synchronic reflections of the life cycle of phonological processes*. University of Manchester PhD thesis.

Wells, John C. 1982. *Accents of English. 3 vols*. Cambridge: Cambridge University Press.

Wrench, Alan A. & James M. Scobbie. 2003. Categorising vocalisation of English /l/ using EPG, EMA and ultrasound. In *Proceedings of the 6th International Seminar on Speech Production*. 314–319.

Wright, Susan. 1987. The interaction of sociolinguistic and phonetically-conditioned CSPs in Cambridge English: auditory and electropalatographic evidence. *Cambridge papers in phonetics and experimental linguistics* 5.

Wright, Susan. 1988. The effects of style and speaking rate on/l/-Vocalisation in Local Cambridge English. *York Papers in Linguistics* 13. 355–365.

Table 1 Summary of fixed effects from a linear model predicting the degree of normalised aperture depending on context and speaker age. The intercept corresponds to a word-medial /l/ pronounced by an older speaker. Positive values indicate increased aperture (less tongue tip raising).

	Dependent variable:
	Aperture (normalised)
Age: younger	−0.013
	(0.066)
Context: word-final pre-C	0.739***
	(0.23)
Context: word-final pre-V	0.177
	(0.173)
Intercept	−0.309**
	(0.13)

Note: *p<0.5; **p<0.01; ***p<0.001

Table 2 Summary of fixed effects from a linear model predicting aperture. The intercept corresponds to /l/ preceded by /u:/ pronounced by an older speaker. Positive values indicate greater aperture (more /l/-vocalisation)

	Dependent variable:
	Aperture (normalised)
Gestural delay	3.806**
	(1.489)
Vowel: ʊ	0.2
	(0.134)
Age: younger	0.024
	(0.231)
F2−F1 distance (centered)	−0.119*
	(0.062)
Intercept	0.497**
	(0.223)

Note: *p<0.5; **p<0.01; ***p<0.001

Table 3 Summary of fixed effects from a linear model predicting the degree of gestural delay. The intercept corresponds to /l/ preceded by /uː/ pronounced by an older speaker. Positive values indicate greater delay.}

	Dependent variable:
	Gestural delay (ms)
Aperture (normalised)	−0.0004
	(0.002)
Vowel: ʊ	−0.05***
	(0.006)
Age: younger	0.012
	(0.033)
F2−F1 distance (centered)	−0.001
	(0.002)
Duration (centered)	−0.039***
	(0.004)
Vowel ʊ: Age younger	0.034***
	(0.007)
Intercept	−0.062***
	(0.023)

Note: *p<0.5; **p<0.01; ***p<0.001